

8-2017

Blending Collaborations and Bridging Gaps: Digital Preservation Communities of Practice - NDSR Lightning Talks

Amy Gay
aegay@binghamton.edu

Meredith Broadway

Joe Carrano

Charlotte Kostelic

Megan Potterbusch

Follow this and additional works at: https://orb.binghamton.edu/librarian_fac



Part of the [Library and Information Science Commons](#)

Recommended Citation

Gay, Amy; Broadway, Meredith; Carrano, Joe; Kostelic, Charlotte; and Potterbusch, Megan, "Blending Collaborations and Bridging Gaps: Digital Preservation Communities of Practice - NDSR Lightning Talks" (2017). *Library Scholarship*. 37.
https://orb.binghamton.edu/librarian_fac/37

This Presentation is brought to you for free and open access by the University Libraries at The Open Repository @ Binghamton (The ORB). It has been accepted for inclusion in Library Scholarship by an authorized administrator of The Open Repository @ Binghamton (The ORB). For more information, please contact ORB@binghamton.edu.



Blending Collaborations and Bridging Gaps: Digital Preservation Communities of Practice

NDSR DC Lightning Talks

Meredith Broadway - World Bank Group

Joe Carrano - Georgetown University

Elizabeth England - Johns Hopkins University

Amy Gay - Food & Drug Administration

Charlotte Kostelic - Library of Congress/Royal Collection Trust

Megan Potterbusch - Association of Research Libraries



Connecting 18th-Century Data for the 21st-Century: George III and George Washington in the Digital Age

Charlotte Kostelic - Library of Congress and Royal
Collection Trust

[@chuckkostelic](https://twitter.com/chuckkostelic)



Background on Georgian Papers Programme

Georgian Papers Programme

What:

- International collaborative project to transform access to the papers of George III and the other Hanoverian monarchs
- Partnership between 7 British and American academic and archival institutions

How:

- Academic exploration
- Technical development
- Collaboration between users and archivists

Why:

- Opportunity to connect early American manuscript collections to their Georgian era counterparts

My Project

What:

- Research the potential for data sharing and interoperability
- Recommend options for future shared access

How:

- Using a comparative analysis
- Researching best practices and data sharing opportunities available
- Developing use cases





What are our current practices?

Current practices

Metadata Quality Measure	Compliance indicator
Completeness	Compliance with GPP metadata application profile/Library of Congress application profile
Accuracy	GPP metadata documentation; known-item search tests; visual view of exported metadata
Conformance to expectations	Development of cataloging practices with users in mind
Consistency and coherence	Visual view; knowledge of data used within research community
Accessibility	Use of standard format and extensible language for metadata creation; use of format and language that is commonly used by community



International
Image
Interoperability
Framework

Level	Name	Contents clearly defined
L6	Conceptual	Documented conceptual model
L5	Dynamic	Effect of information exchanged
L4	Pragmatic	Context of information exchanged
L3	Semantic	Content of information exchanged
L2	Syntactic	Format of information exchanged
L1	Technical	Symbols of information exchanged



How can we connect our collections?

Technical solutions
available





Who are our users? What are their needs?



Name: Drew
Role: Digital Humanist
Motivations: Tenure
Goals: Publish their website
Frustrations: Inconsistent data formatting



Name: Aubrey
Role: Assistant Professor
Motivations: Tenure
Goals: Finish their book
Frustrations: They don't have the time or funding to visit all the archives they need to visit this summer

User needs



Name: Taylor
Role: Teacher
Motivations: Students' learning
Goals: Help students understand history
Frustrations: Too much stuff to sift through



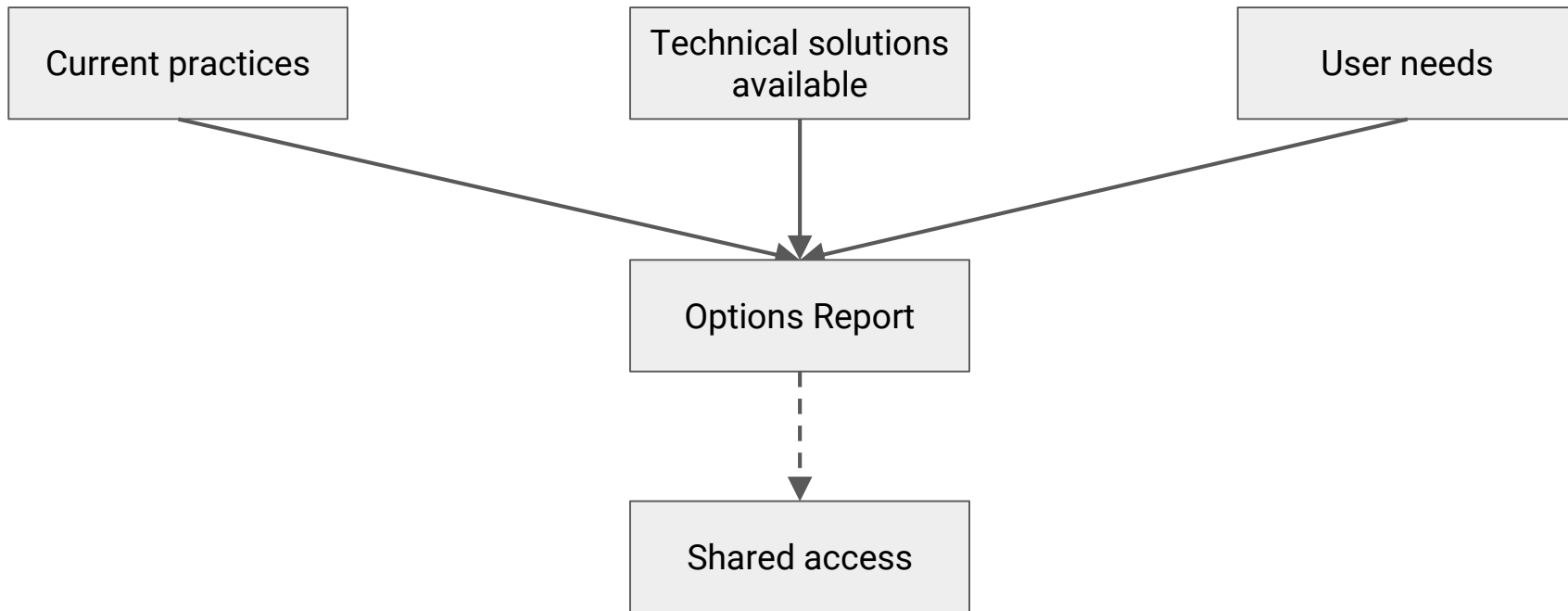
Name: Ari
Role: Archivist
Motivations: Helping users
Goals: Complete reference inquiries
Frustrations: Answering the same questions over and over



Name: Gerry
Role: Generalist
Motivations: Curiosity
Goals: Learn a little bit more about history
Frustrations: The website is confusing



Putting it all together - next steps





Bringing It All Home: Building Digital Preservation Processes for Digital Preservation Platforms

Joe Carrano - Georgetown University
[@joecar25](#)



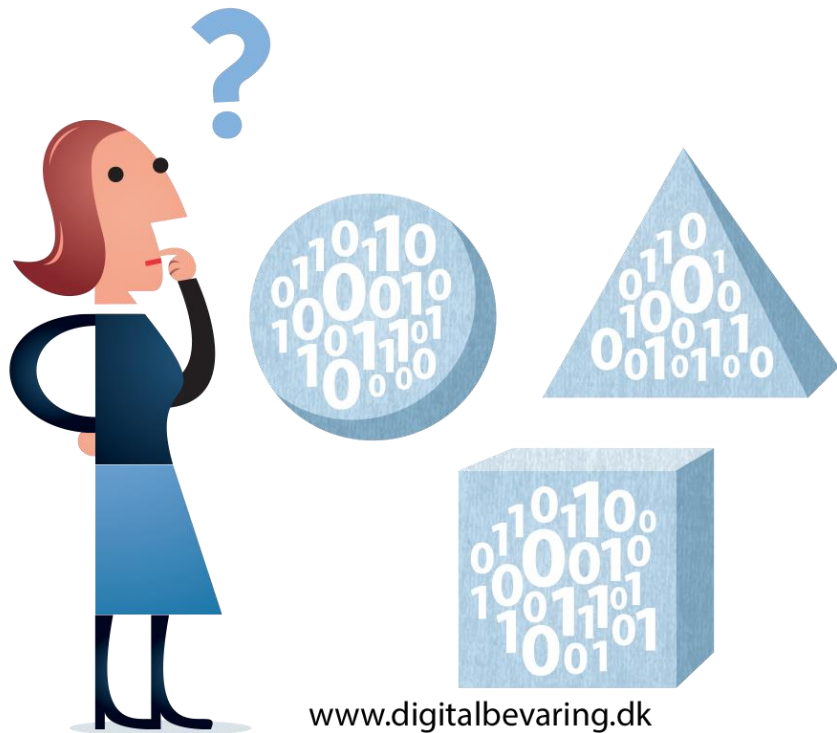
Digital Preservation Foundations

- Georgetown University Library already had a digital preservation plan to manage their digital content
 - Stated goals to build infrastructure, training, and national and international collaboration
- Decided on Academic Preservation Trust (APTrust) as their digital preservation solution
- Built tools for managing and ingesting content
- Needed to develop workflows and implement tools





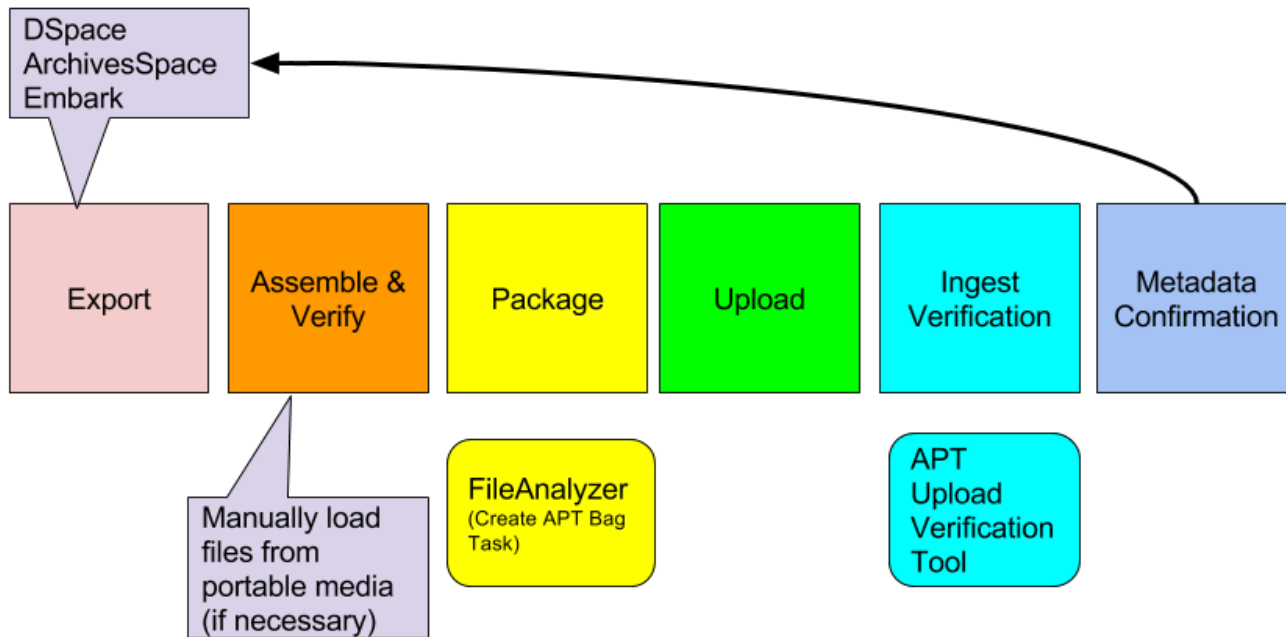
What do we have and where is it?



- Inventoried preservation files that were dispersed throughout the library on different mediums and platforms
- Is this really the preservation copy?
- Cooperation with librarians to determine what happened with legacy projects



Creating workflows and integrations



Some portions of the workflow are open source:

<https://github.com/Georgetown-University-Libraries/File-Analyzer>

<https://github.com/Georgetown-University-Libraries/APTUploadVerification>

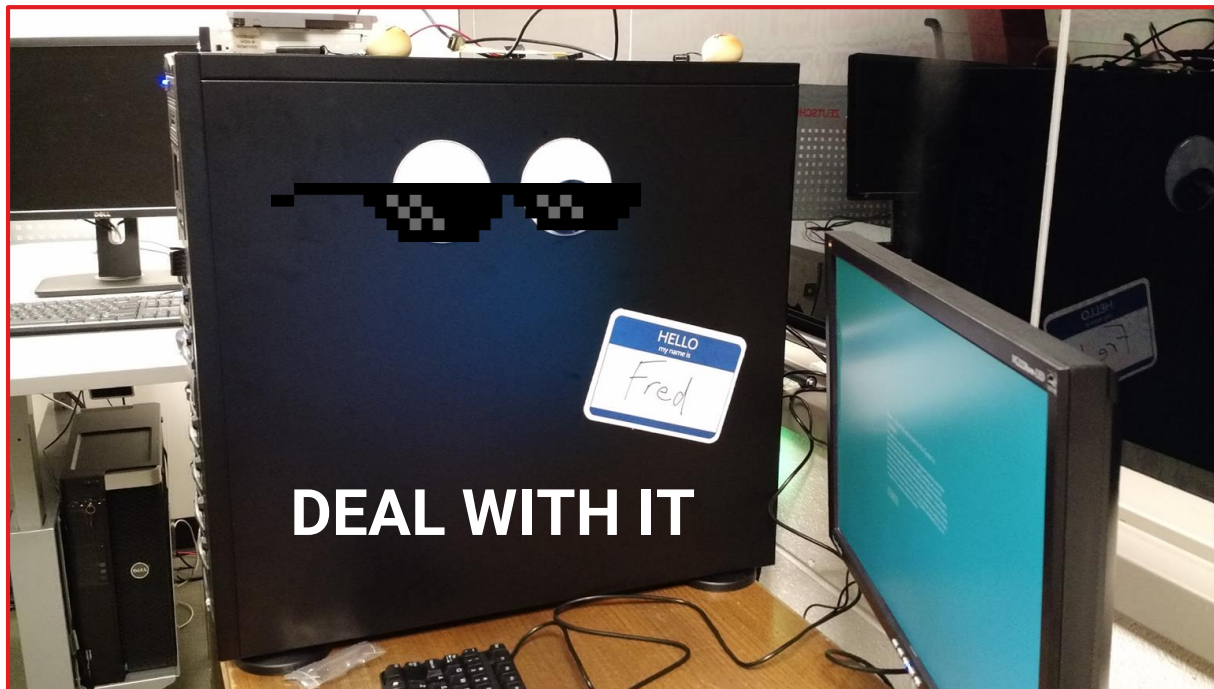


Not all the files were ready to go



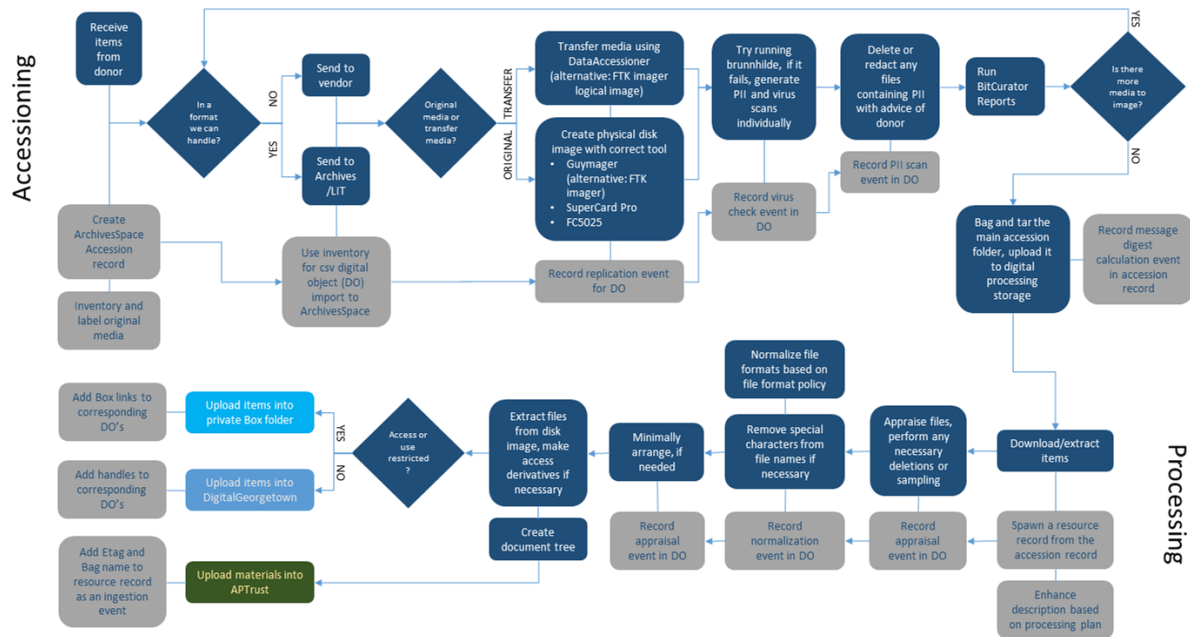
Enter F.R.E.D.

- FRED (Forensic Recovery of Evidence Device)
- Installed and learned more about digital forensics tools like BitCurator
- Further expanded our capacity for moving materials off of obsolete media (floppy disks, zip disks) through purchase of drives and controller cards





Digital Archival Workflows

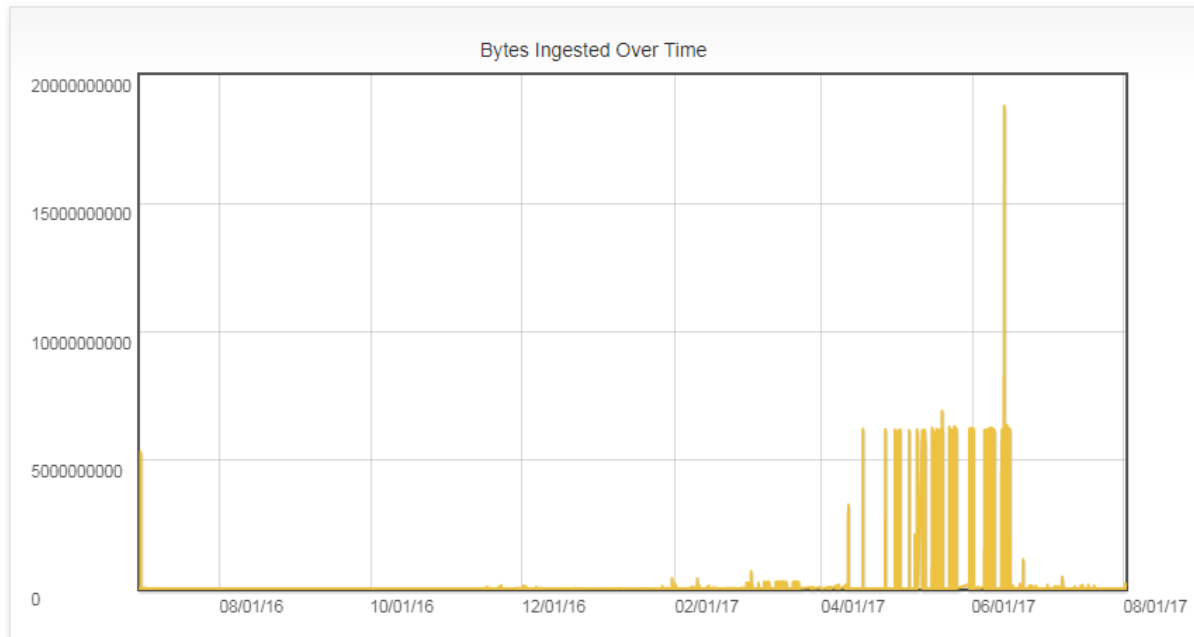


- Collaborated with archivist to create workflow best for them
- Left most traditional archival decisions to the archivists
- Began training others throughout the special collections on how to use FRED and other forensic tools
- Almost all of special collections signed up for training!



Overall Progress

- Ingested over 2.38 TB (and growing) of data into APTrust
- Created ongoing workflows to ingest materials as it's added to the digital access repository
- Trained other librarians on workflows related to their job duties



Graph generated by APTrust. It shows largest file in bytes uploaded per day, not actual number of bytes total ingested per day



Constant Vigilance

When undertaking digital preservation through a third party, it's easy to become complacent that everything is fine

But using these services doesn't mean you should not take the time to check if they're doing what they claim

I discovered a bug in the APTTrust ingest process that caused storage failure and worked with the developer to help fix it

Read more about it here ----->



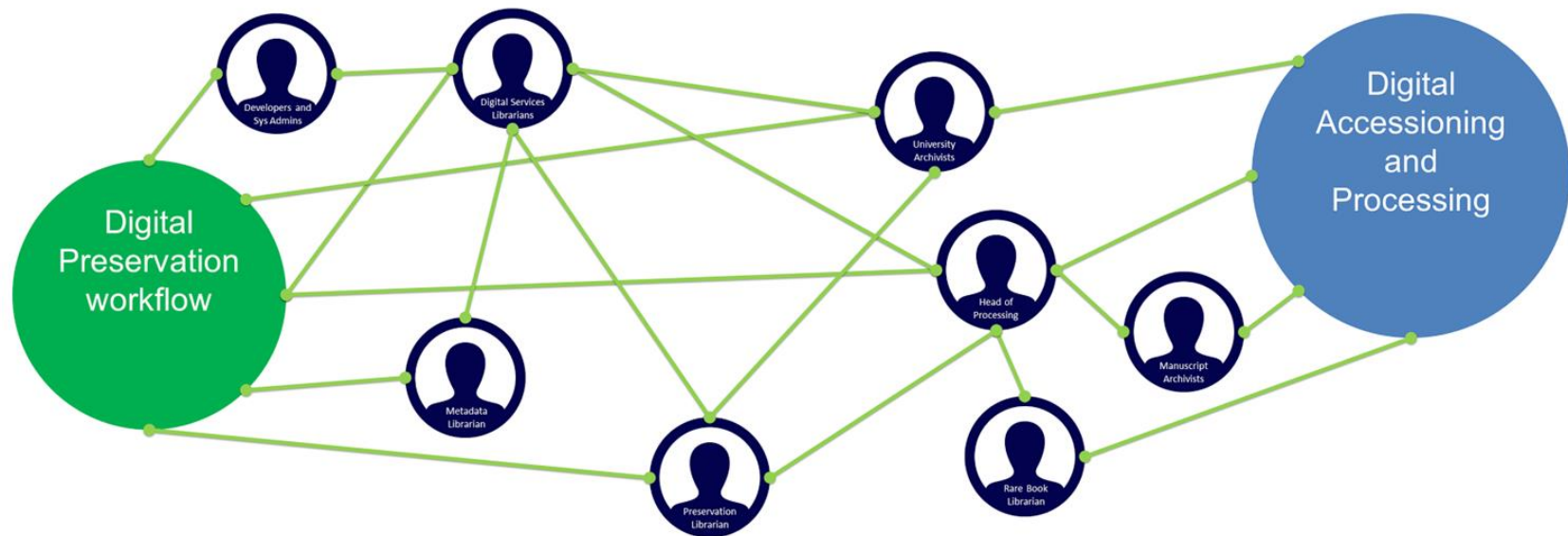
DIGITAL PRESERVATION PRESTIDIGITATION

I realize I haven't posted in awhile but I've been busy, what can I say? I've been continuing to ingest our preservation files into APTTrust, creating workflows for collections with ongoing additions, and integrating our workflows with ArchivesSpace. A couple weeks ago we even passed 1 TB ingested into APTTrust!

<https://josephcarrano.wordpress.com/2017/06/13/digital-preservation-prestidigitation/>



Digital Preservation as a Distributed Responsibility





Data Stewardship and Preservation Program

Meredith Broadway - World Bank Group
[@broadwym](#)



Bridging Units

DECDG (Development Data Group under the Chief Economist)

- Mentor: Malar Veerappan
- Development Data Hub (DDH)
 - Federated repository for Bank's development data
 - Microdata, Time-series, Geospatial
 - Weekly meetings x2

ITSKI (Information Technology Solutions & Knowledge and Information)

- Mentor: Jeanne Kramer-Smyth
- Archives Digital Vault
 - RFI/RFP in-process
 - Call for DDH appraisal guidelines
 - Linking datasets to related documents and reports (D&R external repository)



Guidelines for Appraisal

- WB Archives needs selection criteria because of the volume of DDH data.
- Challenging to apply WB retention schedules to DDH datasets.
- Guidelines require specificity.
 - e.g. Purchasing Power Parities (PPP), “Peace building”, Impact Evaluations
- Developing guidelines:
 - 1. Conducting informal interviews with microdata, geospatial, and time-series leads.
 - 2. Sustained meetings with DDH curators and Archives team.
 - 3. Keeping track of DDH metadata.
- Criteria based on “The UNESCO/PERSIST guidelines for the selection of digital heritage for long-term preservation.”



Guidelines for Appraisal

1. Legal/Due Diligence/Accountability
 - a. How datasets support Bank's mission. Holds institution accountable.
2. Significance
 - a. E.g. Purchasing Power Parities (PPP)
3. Sustainability
 - a. Formats, interoperability, price of dataset, etc.
4. Availability
 - a. Uniqueness; 'Is this dataset duplicated/Does it exist elsewhere?'



Bi-directional linking

- Linking datasets to relevant documents in D&R.
- Direct Object Identifier (DOI), a type of persistent identifier.
- Challenge: hitting a moving target.
- On D&R side, chose to link via ReportID.
- Will eventually be added as a Resource metadata field in DDH.



Site-specific ESIA NG Connection 1.5 Million HHs-Sohag Governorate/ Tahta - September 2016

1 Introduction

1.1 Preamble

The Government of Egypt (GoE) has immediate priorities to increase household use of natural gas (NG) by connecting 1.2 million households/yr to the gas distribution network to replace the highly subsidized, largely imported Liquefied Petroleum Gas (LPG).

The total installed domestic connections until 2015 reached 7.2 million customers and 14.8 thousand commercial customers and 2.3 thousand industrial clients. That was achieved in full cooperation with the local distribution companies.

The GoE is implementing an expansion program for Domestic Natural Gas connections to an additional 1.5 Million households over the next 4 years. The project presented in this study is part of a program that involves extending the network and accompanying infrastructure to connect 1.5 million Households in 11 Governorates between 2016 and 2019 with the assistance of a World Bank Loan of up to US\$500 Million and the Agence Française de Développement (French Agency for Development) financing of up to €70 Million. The program is estimated to cost US\$850 Million.

1.2 Project Objectives

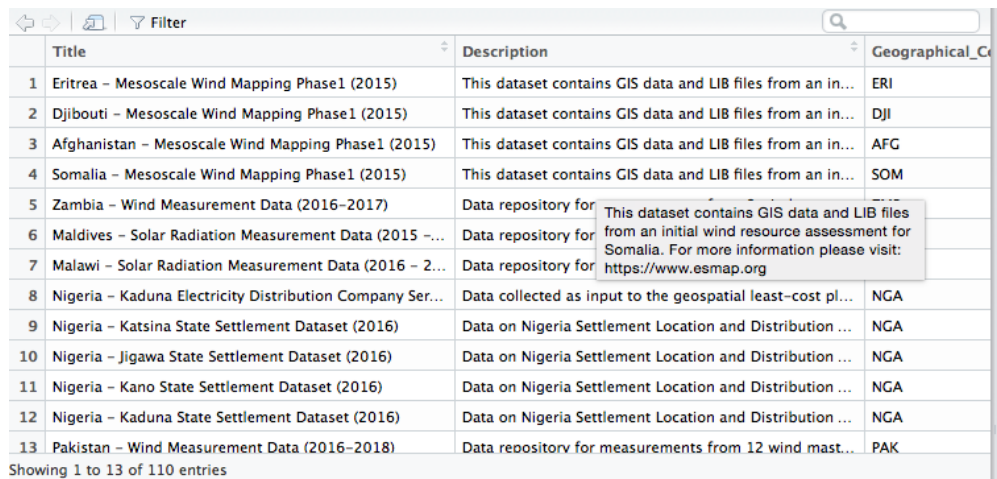
The proposed project represents an integral component of the National energy strategy which aims for greater use of natural gas for domestic users and reduction of government subsidies of the energy sector (LPG). The project is planned for completion within 3 years. The following results are envisaged:

- Wider NG coverage and stable household energy supply



EEX-DDH Harvester

- Given opportunity to help DECDG create a harvester for computationally ingesting energydata.info (EEX) datasets into DDH.
- Challenge: EEX built on CKAN, DDH on DKAN
- Required harmonization of key-value pairs (using list of values and taxonomy), harvesting full representation of data in CKRAN datastore, as well as mapping to DDH-appropriate JSON schema.
- Used R: ckanr, jsonlite, plyr, purrr, tibble



	Title	Description	Geographical_Country
1	Eritrea - Mesoscale Wind Mapping Phase1 (2015)	This dataset contains GIS data and LIB files from an in...	ERI
2	Djibouti - Mesoscale Wind Mapping Phase1 (2015)	This dataset contains GIS data and LIB files from an in...	DJI
3	Afghanistan - Mesoscale Wind Mapping Phase1 (2015)	This dataset contains GIS data and LIB files from an in...	AFG
4	Somalia - Mesoscale Wind Mapping Phase1 (2015)	This dataset contains GIS data and LIB files from an in...	SOM
5	Zambia - Wind Measurement Data (2016-2017)	Data repository for	
6	Maldives - Solar Radiation Measurement Data (2015 - ...)	Data repository for	
7	Malawi - Solar Radiation Measurement Data (2016 - 2...	Data repository for	
8	Nigeria - Kaduna Electricity Distribution Company Ser...	Data collected as input to the geospatial least-cost pl...	NGA
9	Nigeria - Katsina State Settlement Dataset (2016)	Data on Nigeria Settlement Location and Distribution ...	NGA
10	Nigeria - Jigawa State Settlement Dataset (2016)	Data on Nigeria Settlement Location and Distribution ...	NGA
11	Nigeria - Kano State Settlement Dataset (2016)	Data on Nigeria Settlement Location and Distribution ...	NGA
12	Nigeria - Kaduna State Settlement Dataset (2016)	Data on Nigeria Settlement Location and Distribution ...	NGA
13	Pakistan - Wind Measurement Data (2016-2018)	Data repository for measurements from 12 wind mast...	PAK

Showing 1 to 13 of 110 entries



Conclusions

- Attending team meetings ➡ professional growth.
- Sustained engagement between units - key to residency.
- Organization: biweekly meetings with mentors.
- “You don’t know what you don’t know.” Importance of asking questions.
- Moving forward: seeking to develop technical skills.



Bringing Life to Research Objects: Managing the Digital Lifecycle of Research from Creation to Stewardship through the Open Science Framework (OSF) and SHARE

Megan Potterbusch - Association of Research Libraries

[@LibrarPotter](#)



Digital Stewardship Starts Early

“Stewardship is an ethic that embodies the responsible planning and management of resources.”

- The Open Science Framework (OSF) is designed to support entire research lifecycle
- “Registrations” feature bundles workspace to “respect des fonds” of research





Contributors to My Project

- Association for Research Libraries
 - Judy Ruttenberg
 - Prue Adler
- George Washington University
 - Hannah Sommers
 - Bill Gillis
 - Mandy Gooch (now at University of North Carolina at Chapel Hill)
- Center for Open Science
 - Jeff Spies
 - Courtney Soderberg
 - Natalie Meyers (also at University Notre Dame)





Project Goals

- Work with research units to implement OSF as a step towards digital stewardship
- Building resources to support new library services
- Outreach to other librarians (at GW and elsewhere) about this work



Research Teams & Example Needs

- Humanities
 - New project management tool
 - Optional public facing component
- Life Sciences
 - Public landing site with added information
 - Tool accessible to agencies blocked from other services
- Interdisciplinary Science
 - Dispersed team management
 - Exploring the disruptive potentials of new technology



Outreach to Librarians

Presentations, workshops, and trainings locally and at conferences



code{4}lib





Watch This Space

- “OSF Toolkit for Digital Scholarship Support”
 - <https://osf.io/pdsmb/> (Currently private but link will work when public)
- Fedora integration with OSF (currently in beta)
- Open Science Framework Train the Trainer (<https://osf.io/adkj4/>)



Enabling Open Science through the Center for Devices & Radiological Health (CDRH) Science Data Catalog

Amy Gay - Food & Drug Administration
@iamabooknerd

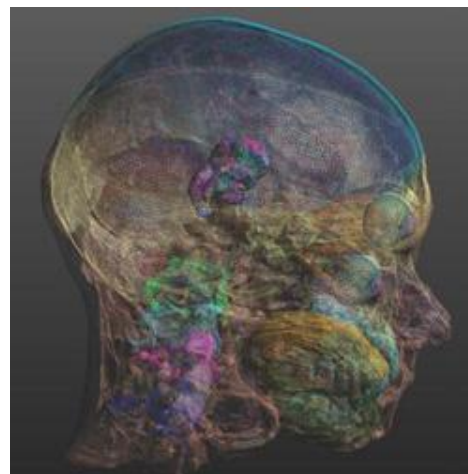
Mentors: Jessica Berrellez & Anindita Saha

*This presentation reflects the views of the author and should not be construed to represent
FDA's views or policies*



Background on U.S. FDA's CDRH & OSEL

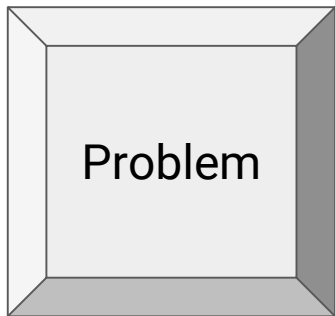
Center for Devices & Radiological Health
Office of Science & Engineering Laboratories
Mission: To protect and promote Public Health



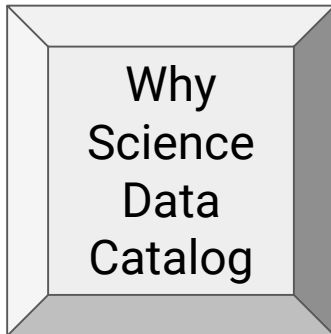
Images from
<https://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHOices/ucm115989.htm>



Pilot Project



- Federally mandated public access policies
- No institutional repository
- No current tracking method



- One-stop shop for datasets and publications
- Compliance with public access policies



- Internal & Public Facing Catalog
- User Testing
- Workflows
- Formal Presentations to Stakeholders

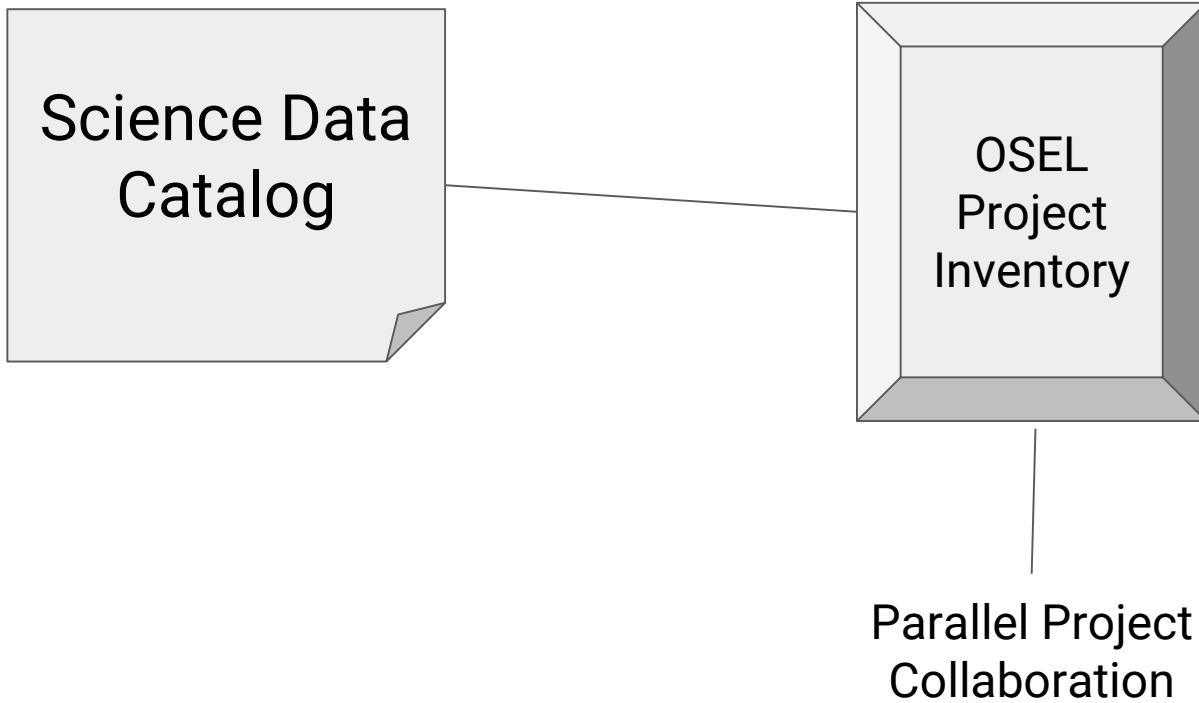


Web of Collaboration

Science Data
Catalog

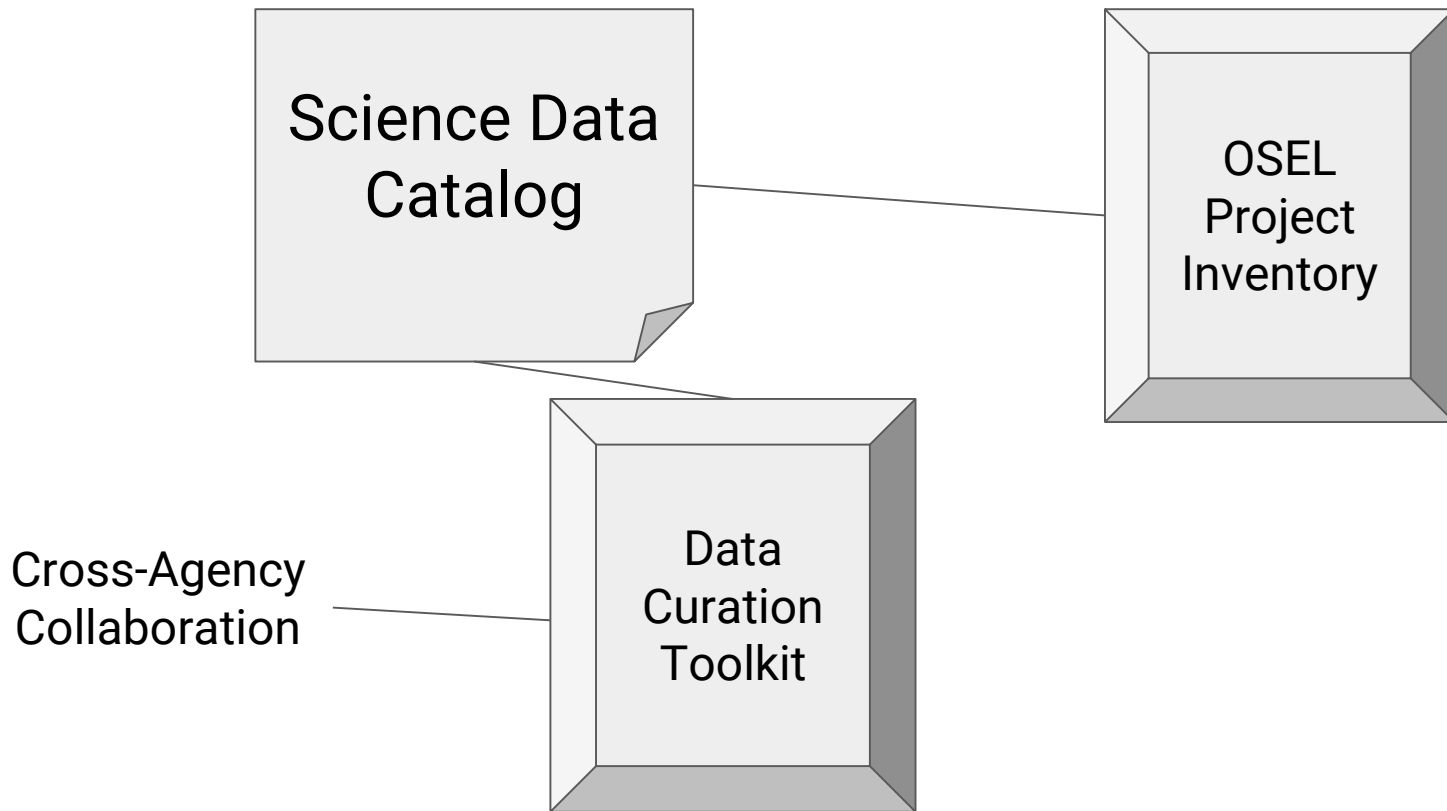


Web of Collaboration



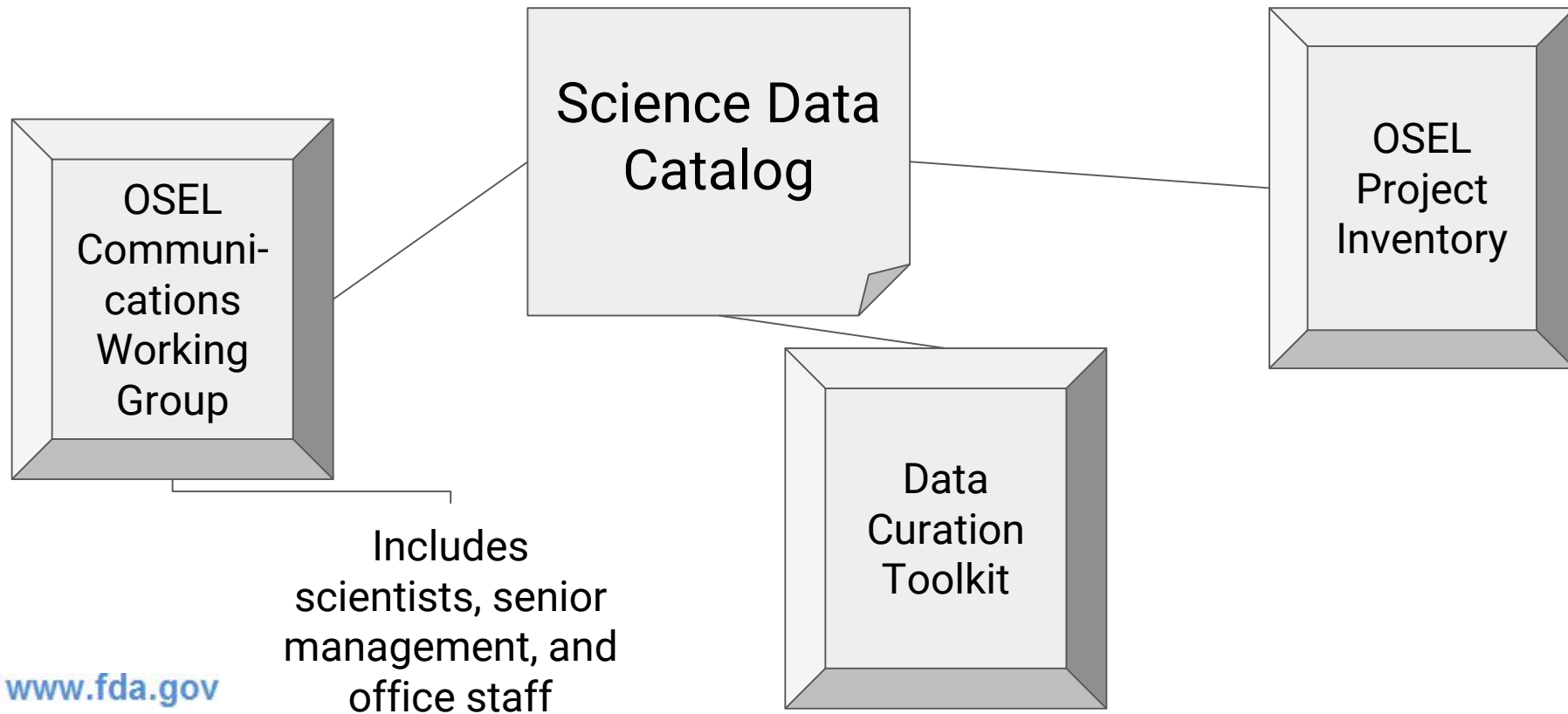


Web of Collaboration



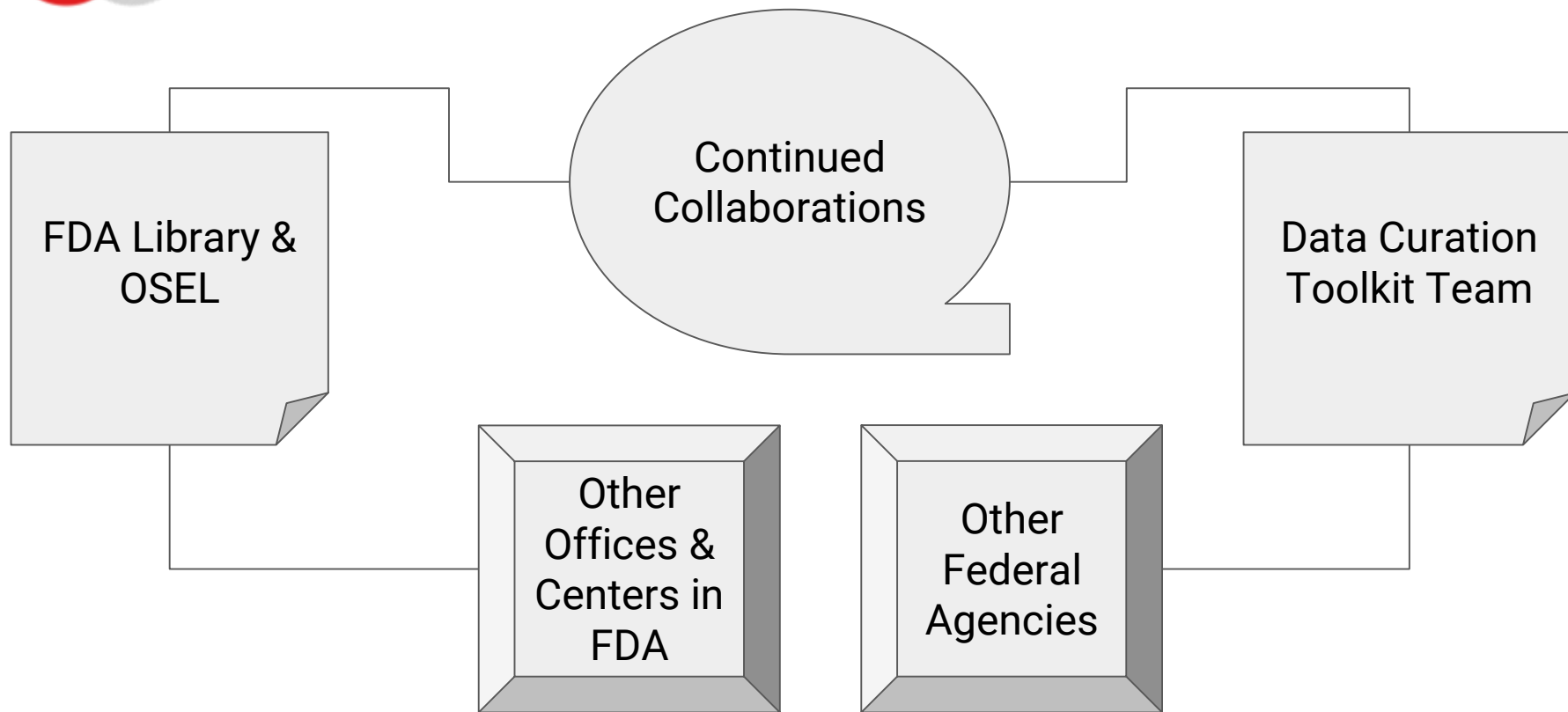


Web of Collaboration





Future / Post Residency





THANK YOU!